Available online at www.sciencedirect.com

ScienceDirect

Journal of Multivariate Analysis 98 (2007) 256–281

Journal of Multivariate Analysis

www.elsevier.com/locate/jmva

A test for elliptical symmetry

Fred W. Huffer[a,*], Cheolyong Park[b]

[a] *Department of Statistics, Florida State University, Tallahassee, Florida 32306-4330, USA*
[b] *Department of Statistics, Keimyung University, Taegu 704-701, Korea*

Received 19 February 2005
Available online 8 November 2005

Abstract

This paper presents a statistic for testing the hypothesis of elliptical symmetry. The statistic also provides a specialized test of multivariate normality. We obtain the asymptotic distribution of this statistic under the null hypothesis of multivariate normality, and give a bootstrapping procedure for approximating the null distribution of the statistic under an arbitrary elliptically symmetric distribution. We present simulation results to examine the accuracy of the asymptotic distribution and the performance of the bootstrapping procedure. Finally, for selected alternatives, we compare the power of our test statistic with that of recently proposed tests for elliptical symmetry given by Manzotti et al. [A statistic for testing the null hypothesis of elliptical symmetry, J. Multivariate Anal. 81 (2002) 274–285] and Schott [Testing for elliptical symmetry in covariance-matrix-based analyses, Statist. Probab. Lett. 60 (2002) 395–404], and with that of the well known tests for multivariate normality of Mardia [Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530] and Baringhaus and Henze [A consistent test for multivariate normality based on the empirical characteristic function, Metrika 35 (1988) 339–348].

AMS 1991 subject classification: 62H15; 62E20

Keywords: Elliptically contoured distribution; Multivariate normality; Chi-square test; Bootstrap approximation

1. Introduction

The family of elliptically symmetric (or elliptically contoured) distributions is an important class of distributions which generalizes the family of multivariate normal distributions. There is a large literature on these distributions and their use in statistics (see [7,8]). The assumption of

* Corresponding author.
E-mail addresses: huffer@stat.fsu.edu (F.W. Huffer), cypark1@kmu.ac.kr (C. Park).

0047-259X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2005.09.011

elliptic symmetry plays an important role in applications; for example, in sliced inverse regression [16], and in the distribution of asset returns [11].

This paper introduces a new test for elliptical symmetry which is now briefly described. Suppose we have data $y_1, y_2, \ldots, y_n$ consisting of random $p \times 1$ vectors which are iid from a density f on $\mathbb{R}^p$. The statistic we propose for testing the elliptical symmetry of f is a Pearson's chi-square statistic. We divide the space $\mathbb{R}^p$ into cells which have (asymptotically) equal expected cell counts under the hypothesis of elliptical symmetry. We determine the observed counts for these cells, and then compute the chi-square statistic X^2 for comparing the observed and expected cell counts.

The statistic X^2 is most easily described and computed in terms of the scaled residuals $z_1, z_2, \ldots, z_n$ defined as follows. Let $\bar{y}$ and S be the sample mean vector and covariance matrix of $y_1, \ldots, y_n$ given by $\bar{y} = n^{-1} \sum_i y_i$ and $S = n^{-1} \sum_i (y_i - \bar{y})(y_i - \bar{y})^t$. Define

$$z_i = R(y_i - \bar{y}) \quad \text{for } i = 1, \ldots, n, \tag{1.1}$$

where the matrix $R = R(S)$ is a function of S satisfying $RSR^t = I$. The coordinates of the scaled residuals have sample correlations of zero, and are standardized to have sample means and variances equal to zero and one, respectively. The scaled residuals are also known as the spherized data, and the process of computing the scaled residuals will be called "spherizing" the data. There are various ways of spherizing the data in common use. A principal components transformation of the data corresponds to choosing a particular matrix R of the form $D\Gamma$ where Γ is an orthogonal matrix and D is a diagonal matrix. A Gram–Schmidt transformation takes R to be lower triangular with positive diagonal elements. Another commonly used transformation uses $R = S^{-1/2}$.

In computing X^2, the space of the scaled residuals is divided into cells. First, divide $\mathbb{R}^p$ into c spherical shells centered at the origin, with each shell containing an equal number of the scaled residuals z_i. That is, the $c - 1$ radii of the boundaries between the shells are the j/cth sample quantiles, $j = 1, \ldots, c-1$, of the lengths $|z_1|, |z_2|, \ldots, |z_n|$. Secondly, divide $\mathbb{R}^p$ into g congruent sectors emanating from the origin. The sectors are to be congruent in the sense that, for any pair of sectors, there is an orthogonal transformation mapping one into the other. (For example, the 2^p orthants give one possible set of sectors. In $\mathbb{R}^2$, sectors are just equiangular slices or wedges.) Taken together, the c shells and g sectors divide $\mathbb{R}^p$ into gc cells which (at least asymptotically) are each expected to contain $n/(gc)$ of the vectors z_i. We compute the observed counts (the number of z_i) in each cell, and our statistic X^2 is the chi-square statistic comparing these observed counts with the values $n/(gc)$ expected under elliptical symmetry.

Our approach to testing elliptical symmetry has its roots in the work of Moore and Stubblebine [20] who devised a chi-square test for multivariate normality based on the use of spherical shells.

If we knew $\mu = Ey_i$ and $\Sigma = \mathrm{Var}(y_i)$, we could use these in place of $\bar{y}$ and S in computing the scaled residuals in (1.1), and then under the hypothesis of elliptical symmetry $z_1, \ldots, z_n$ would be iid from a spherically symmetric distribution. In this case, it is easily seen that the cell counts in the different shells are independent, and that, within each of the c shells, the counts in the g cells follow a multinomial distribution. Thus, the classical result of Pearson applies within each shell to say that the contribution to X^2 from the terms in each shell has an asymptotic χ^2_{g-1} distribution, and pooling these over the c shells gives an asymptotic $\chi^2_{c(g-1)}$ distribution for X^2. Thus, when the spherizing is done using the true values of μ and Σ, the distribution of X^2 is the same for all elliptically symmetrical distributions, and the asymptotic distribution is a simple chi-squared.

Unfortunately, when the spherizing is done using $\bar{x}$ and S, the situation is more complicated. The null distribution of X^2 now depends somewhat on the underlying elliptically symmetric distribution. It also depends, when $p \geqslant 3$, not just on the values of c and g, but also on the particular method of dividing $\mathbb{R}^p$ into sectors. These complications will occupy much of this paper.

Our paper is organized as follows. In Section 2 we present some background material on elliptically symmetric distributions, give a more detailed description of our statistic X^2, and then present our main results. In particular, we give the general form of the asymptotic distribution of our statistic X^2 under the assumption of multivariate normality and describe a bootstrapping procedure for approximating the distribution of our statistic under an arbitrary elliptically symmetric distribution. In Section 3 we describe some other tests for elliptical symmetry, including the statistics of Manzotti et al. [17] and Schott [23] which we use in our later comparisons. Sections 4 and 5 present simulation results which study the accuracy of the asymptotic distribution for finite sample sizes and the performance of the bootstrap procedure, and also compare the power of our procedure with other tests for elliptical symmetry and multivariate normality under various alternatives. Section 6 contains a proof of the asymptotic distribution of X^2. This asymptotic distribution is a linear combination of chi-squared random variables, and Sections 7 and 8 describe how to compute the coefficients (eigenvalues) involved in this linear combination for different types of sectors.

2. Definitions and main results

The random $p \times 1$ vector y has a spherically symmetric distribution if

$$y \overset{\mathrm{d}}{=} \xi U, \tag{2.1}$$

where U is uniformly distributed on the surface of the unit sphere in $\mathbb{R}^p$ (henceforth denoted by $\Omega_p = \{x \in \mathbb{R}^p : |x| = 1\}$), ξ is a random nonnegative scalar, and U and ξ are independent. Equivalently, y is spherically symmetric if $y \overset{\mathrm{d}}{=} \Gamma y$ for all orthogonal matrices Γ. An affine transformation of a vector with a spherically symmetric distribution produces a vector with an elliptically symmetric distribution: y has an elliptically symmetric distribution if

$$y \overset{\mathrm{d}}{=} \mu + \xi A U \tag{2.2}$$

for a $p \times p$ nonsingular matrix A and $\mu \in \mathbb{R}^p$ where U and ξ are as in (2.1). It is easily seen that the distribution of y depends on A only through $\Sigma = AA^t$, and that, if $E\xi^2 < \infty$, then $Ey = \mu$ and $\mathrm{Var}(y) = (p^{-1}E\xi^2)\Sigma$. Throughout this paper we shall assume $E\xi^2 < \infty$. Then without loss of generality we may take $E\xi^2 = p$ so that $\mathrm{Var}(y) = \Sigma$. An elliptically symmetric distribution is determined by the parameters $\theta = (\mu, \Sigma)$ and the distribution of ξ. We shall also assume throughout that our elliptically symmetric distributions have densities $f_\theta(y)$ which must necessarily have the form

$$f_\theta(y) = |\Sigma|^{-1/2} h\left((y-\mu)^t \Sigma^{-1} (y-\mu)\right)$$

for some function h.

In computing our statistic X^2, we must choose a method for spherizing the data. Throughout this paper, we will use the Gram–Schmidt transformation because, if the data $y_1, \ldots, y_n$ are iid from an elliptically symmetric distribution, the Gram–Schmidt transformation leads to scaled residuals $z_1, \ldots, z_n$ in (1.1) which are ancillary, that is, their joint distribution does not depend on μ or Σ (see Lemma 6.1). For nonnormal data, simulations show that both the principal components transformation and $R(S) = S^{-1/2}$ lead to scaled residuals whose joint distribution depends on

Σ. However, this dependence is typically negligible except in small samples. (For multivariate normal data, Lemma 3.1 in Huffer and Park [12] shows that the scaled residuals are ancillary for all choices of $R(S)$, and moreover, the joint distribution of $z_1, \ldots, z_n$ is the same for all choices of $R(S)$.) Since we will use Gram–Schmidt, our statistic X^2, which is a function of the scaled residuals, is also ancillary, so that we may take $\mu = 0$ and $\Sigma = I$ in all our later simulations.

We now describe the construction of the cells for our chi-square statistic. Let $g \geqslant 2$ and $c \geqslant 1$ be integers. In two dimensions, the g sectors used in constructing our cells are easily described and visualized. But for a general discussion of the situation in dimension three or more, it is most convenient to describe the sectors as congruent regions which are generated by applying a group of orthogonal transformations to a given fundamental region G.

Let $\mathcal{G} = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_g\}$ be a finite group of orthogonal $p \times p$ matrices with $\Gamma_1 = I$. Let $G \subset \mathbb{R}^p$ be a region which satisfies:

(1) G is open;
(2) $G \cap \Gamma_i G = \emptyset$ for $i \neq 1$;
(3) $\mathbb{R}^p = \cup_{i=1}^{g} \overline{\Gamma_i G}$ (here $\overline{A}$ denotes the closure of A);
(4) G is an intersection of finitely many half-spaces: there exist vectors $a_1, a_2, \ldots, a_k$ such that $G = \{x : a_i^t x > 0 \text{ for } i = 1, \ldots, k\}$.

Clearly (4) implies (1). Note that (4) implies that if $x \in G$, then $\lambda x \in G$ for all $\lambda > 0$, $\lambda \in R$. A region satisfying (1)–(3) is called a fundamental region for the group $\mathcal{G}$. For every finite group $\mathcal{G}$ there exists a region G satisfying (1)–(4). See [9, Chapter 3]. Throughout the following, let G denote such a region.

Define the sectors G_i by $G_i = \Gamma_i G$ for $i = 1, 2, \ldots, g$. The c shells are defined as follows. Let $\hat{F}_n$ be the empirical cdf of the values $\{|z_i|^2 : 1 \leqslant i \leqslant n\}$, and define the sample quantiles $q_{nj} = \hat{F}_n^{-1}(j/c)$ for $j = 1, \ldots, c-1$. For $1 \leqslant j \leqslant c$ define the shell S_{nj} by

$$S_{nj} = \{z : q_{n\,j-1} < |z|^2 \leqslant q_{nj}\},$$

where we take $q_{n0} = 0$ and $q_{nc} = \infty$.

We divide the space of the scaled residuals into gc cells $\Lambda_{n\pi}$ as follows. For $\pi = (i, j)$ with $1 \leqslant i \leqslant g$ and $1 \leqslant j \leqslant c$, define

$$\Lambda_{n\pi} = G_i \cap S_{nj}.$$

The cell counts $U_{n\pi}$ are given by

$$U_{n\pi} = \sum_{k=1}^{n} I(z_k \in \Lambda_{n\pi}) \tag{2.3}$$

and in terms of them the chi-square statistic X^2 is

$$X^2 = \sum_{\pi} (U_{n\pi} - np_0)^2 / np_0, \tag{2.4}$$

where $p_0 = 1/(gc)$.

In Section 6 we derive the limiting distribution of X^2 when the data $y_1, \ldots, y_n$ are iid from a multivariate normal distribution. The limiting distribution is a linear combination of chi-squared random variables:

$$X^2 \xrightarrow{\text{d}} W_0 + \sum_{i=1}^{m} (1 - \lambda_i) W_i \quad \text{as } n \to \infty, \tag{2.5}$$

where $W_0, W_1, \ldots, W_m$ are independent, $W_0 \sim \chi^2_{c(g-1)-m}$, $W_1, \ldots, W_m$ are iid χ^2_1, and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the nonzero eigenvalues of QQ^t with the matrix Q being defined in Eq. (6.7). We give a recipe in Section 7 for the calculation of these eigenvalues and use this in Section 8 to compute the eigenvalues for a number of different configurations of cells arising from different choices of the group $\mathcal{G}$ and fundamental region G. Our simulation results indicate that the limiting distribution is a good approximation to the distribution of X^2 so long as the average number of observations per cell $n/(gc)$ is not too small; $n/(gc) \geqslant 5$ is usually more than adequate.

When the data is sampled from an elliptically symmetric distribution which is not normal, the simulations in Section 4 suggest that the limiting distribution in (2.5) is still a reasonable approximation to the distribution of X^2 so long as the distribution we are sampling from is not too far removed from the normal distribution. For distributions which differ greatly from the normal, a bootstrapping procedure like that in Koltchinskii and Sakhanenko [15] can be used to approximate the distribution of X^2. This bootstrapping procedure uses the empirical distribution of the lengths $\xi_i = |z_i|$, $i = 1, \ldots, n$ to approximate the distribution of ξ in (2.2), and consists of repeating the following steps:

1. Sample $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ with replacement from the values $\xi_1, \xi_2, \ldots, \xi_n$.
2. Generate $U_1^*, U_2^*, \ldots, U_n^*$ which are iid uniform on Ω_p.
3. Compute $y_i^* = \xi_i^* U_i^*$ for $i = 1, \ldots, n$.
4. Treat this as "raw" data and use it to compute the corresponding chi-square value X^{2*}.

Doing this many (say, J) times gives a sample of bootstrap replicates $X_1^{2*}, X_2^{2*}, \ldots, X_J^{2*}$ whose distribution approximates the distribution of X^2. In our simulations (see Section 4), this procedure has performed very well. Note that, since we are using the Gram–Schmidt transformation, the distribution of X^2 does not depend on μ and A in (2.2) so that in step 3 above we can take $\mu = 0$ and $A = I$.

3. Some other tests for elliptical symmetry

We shall be comparing the performance of X^2 with two other tests for elliptical symmetry found in the literature. Manzotti et al. [17] develop a test statistic based on the fact that, under the assumption of elliptical symmetry, the normalized vectors $u_i = z_i/|z_i|$, $i = 1, \ldots, n$ will be asymptotically uniformly distributed on Ω_p. They test for nonuniformity using a statistic based on spherical harmonic functions (see [18,4]). Let $\mathcal{H}_i$ denote an orthonormal basis for the set of spherical harmonics of degree i, and $\mathcal{J}_{jl} = \cup_{j \leqslant i \leqslant l} \mathcal{H}_i$. For a given spherical harmonic function h, the deviation from nonuniformity in the "direction" of h is measured by

$$Q_n(h) = \frac{1}{n} \sum_{i=1}^{n} h(u_i) I\left(|z_i|^2 > \hat{F}_n^{-1}(\varepsilon)\right). \tag{3.1}$$

The indicator function deletes the fraction ε of the data having the smallest radii $|z_i|$ (where ε is a small positive value). This is done for technical reasons. Manzotti et al. combine these individual deviations from nonuniformity to form a statistic which we shall refer to by the initials of the authors:

$$MPQ = n \sum_{h \in \mathcal{J}_{jl}} Q_n^2(h), \tag{3.2}$$

where $j \geqslant 3$. They show that the asymptotic distribution of this statistic is $(1-\varepsilon)\chi_v^2$ where v is the number of functions in $\mathcal{J}_{jl}$.

Schott [23] proposes a test (denoted T_1) for elliptical symmetry based on fourth moments. The notation needed to precisely define this statistic is somewhat complicated, but the basic idea is easily expressed: Schott uses a Wald-type test to compare the sample fourth moments of $z_1, \ldots, z_n$ with those expected for an elliptically symmetric distribution. His procedure requires consistent estimates of the covariance matrix of the sample fourth moments, and thus requires the elliptically symmetric distribution which is being sampled from to have finite moments up to order eight. Under this condition, Schott [23, p. 399] shows that his statistic has an asymptotic χ_v^2 distribution with degrees of freedom v depending on the dimension p of the data.

Beran [3] defines a class of tests for elliptical symmetry. His tests are based on scaled residuals, but he allows for a somewhat more general form of spherizing than we consider. For simplicity, we describe Beran's tests in terms of our scaled residuals $z_1, \ldots, z_n$. For $i = 1, \ldots, n$, let r_i be the ranks, divided by $n+1$, of the lengths $|z_i|$, and u_i be the direction vectors $u_i = z_i/|z_i|$. Let $\{a_k : k \geqslant 1\}$ be a family of functions orthonormal with respect to Lebesgue measure on [0, 1] and orthogonal to the constant function on [0, 1]. Let $\{b_m : m \geqslant 1\}$ be a family of functions orthonormal with respect to the uniform distribution on Ω_p and orthogonal to the constant function on Ω_p. Beran's test statistics have the form

$$S_n = \sum_{k=1}^{K_n} \sum_{m=1}^{M_n} \left[n^{-1/2} \sum_{i=1}^{n} a_k(r_i) b_m(u_i) \right]^2 .$$

A very general class of tests has been proposed by Koltchinskii and Sakhanenko [15]. For any function f on $\mathbb{R}^p$, let $m_f(\rho)$ denote the average value of f on the sphere of radius ρ about the origin, that is, $m_f(\rho) = Ef(\rho U)$ where U is uniformly distributed on Ω_p. Let $\mathcal{F}$ be a class of functions from $\mathbb{R}^p$ to $\mathbb{R}$ which is closed under orthogonal transformations: if $f \in \mathcal{F}$, then $f \circ \Gamma \in \mathcal{F}$ for any orthogonal transformation Γ. Koltchinskii and Sakhanenko consider test statistics of the form

$$\sup_{f \in \mathcal{F}} n^{-1/2} \sum_{i=1}^{n} \left(f(z_i) - m_f(|z_i|) \right) .$$

The test MPQ and Schott's T_1 are not intended to be omnibus tests of elliptical symmetry, that is, they will not have power against all alternatives. Schott's test will have no (or little) power against alternatives whose fourth moments exactly (or approximately) match those of an elliptically symmetric distribution. The test based on MPQ will have no (or little) power against any alternative for which the direction vectors $u_i = z_i/|z_i|$ are exactly (or nearly) uniformly distributed on Ω_p. Also, in order to ensure that their test has the same asymptotic distribution for all elliptically symmetric distributions, they include only spherical harmonics h of degree greater than 2 in (3.2). As a consequence, MPQ has no power against alternatives where the deviation from uniformity on Ω_p is in the direction of a second degree spherical harmonic. However, the tests of Beran and of Koltchinskii and Sakhanenko can, in principle, be used as omnibus tests. (To achieve this, in Beran's test the values M_n and K_n must go to infinity at an appropriate rate as $n \to \infty$, and in Koltchinskii and Sakhanenko's test the class $\mathcal{F}$ must be chosen appropriately.)

The tests of Manzotti et al., Schott, and Koltchinskii and Sakhanenko all have the desirable property of affine invariance: if the data y_i are replaced by $y_i^* = Ay_i + b$ (where A is nonsingular) for $i = 1, \ldots, n$, the value of the test statistic is not changed. Beran's test will not typically possess this property.

The tests of Manzotti et al., and Schott have a further desirable property: they have a simple asymptotic null distribution which is the same for all elliptically symmetric populations possessing the necessary number of moments (finite fourth order moments for Manzotti et al., and finite eighth order moments for Schott). The tests of Koltchinskii and Sakhanenko do not have a known asymptotic distribution (even when sampling from the multivariate normal distribution); a bootstrapping procedure is needed to obtain the necessary critical values. Also, the asymptotic distribution depends on the particular elliptically symmetric population being sampled from. Beran [3] showed, if $K_n, M_n \to \infty$ appropriately as $n \to \infty$, that $(K_n M_n)^{-1/2}(S_n - K_n M_n)$ has a $N(0, 2)$ limiting distribution for any elliptically symmetric population. But in a few experiments we conducted, for the particular basis functions and values of n, K_n, M_n we used, the $N(0, 2)$ approximation was not accurate enough to supply useful critical points.

In our simulation studies in Sections 4 and 5, we restrict ourselves to the tests of Manzotti et al., and Schott (which we refer to as *MPQ* and *Schott*). The tests of Koltchinskii and Sakhanenko are difficult to compute in addition to having intractable null distributions. For Beran's statistic, there is no guidance in the literature for the choice of basis functions, or the values K_n and M_n, and, as noted above, the asymptotic $N(0, 2)$ distribution is not sufficiently accurate. For these reasons, we did not include these statistics in our simulation studies. Zhu and Neuhaus [25] have recently proposed a new test for elliptical symmetry based on the empirical characteristic function. We have not studied this test, but it appears promising.

The statistics X^2, MPQ and Schott are intended primarily as tests for elliptical symmetry, but we also wish to consider their possible use in testing the more narrow hypothesis of multivariate normality. For this reason we have included some tests for multivariate normality in our simulation studies. A very large number of tests for multivariate normality have been proposed in the literature. In our comparisons, we will include only three: the well-known skewness and kurtosis tests of Mardia [19], and the test statistic T_n of Baringhaus and Henze [1]. We refer to these as *Skew*, *Kurt*, and *BH* in our tables. The choice of these three tests is somewhat arbitrary, but we note that they are all based on different principles, and have performed well in previous studies such as those of Romeu and Ozturk [22] and Manzotti and Quiroz [18]. Mardia's tests are specialized tests designed to detect multivariate skewness and kurtosis, respectively, whereas T_n of Baringhaus and Henze gives an omnibus test with power approaching 1 as $n \to \infty$ for any fixed alternative. The Skew and Kurt statistics have limiting chi-squared and normal distributions, respectively, given in (2.26) and (3.20) of Mardia [19]. The limiting distribution of T_n is not known in any convenient form, but Henze and Wagner [10] give an accurate lognormal approximation (denoted $q_{\beta,d}(\alpha)$).

4. The null distribution of X^2

In this section, we present simulation results illustrating the accuracy of the limiting distribution of X^2 in (2.5) for finite samples from a variety of elliptically symmetric distributions.

A convenient family of elliptically symmetric distributions is the multivariate generalized Laplace distribution, denoted by MGL, suggested by Ernst [6]. The density function of MGL(λ),

Table 1
Simulation results for MGL(2), the multivariate normal distribution

n	Method	$p=3$			$p=4$		
		$\alpha=.01$	$\alpha=.05$	$\alpha=.1$	$\alpha=.01$	$\alpha=.05$	$\alpha=.1$
100	X^2	.00958	.04598	.09670	.00906	.04626	.09860
	MPQ	.00836	.04698	.09730	.00958	.04940	.10068
	Schott	.00566	.03902	.08818	.00710	.04164	.09270
	BH	.00806	.04614	.09262	.00942	.04554	.09250
	Skew	.00970	.04060	.07948	.00970	.03836	.07458
	Kurt	.00580	.02688	.06522	.00456	.02768	.07108
200	X^2	.00932	.05002	.09786	.00960	.05064	.10154
	MPQ	.00924	.04964	.09982	.00988	.04994	.10036
	Schott	.00732	.04170	.09052	.00730	.04344	.09378
	BH	.00962	.04812	.09740	.01022	.04904	.09544
	Skew	.01088	.04640	.09174	.01008	.04538	.08760
	Kurt	.00786	.03666	.08124	.00710	.03856	.08616

indexed by $\lambda > 0$, is given by

$$f(x) = \frac{\lambda\Gamma(p/2)}{2\pi^{p/2}\Gamma(p/\lambda)|\Sigma|^{1/2}} \exp\left\{-[(x-\mu)^t\Sigma^{-1}(x-\mu)]^{\lambda/2}\right\}.$$

Setting $\lambda = 2$ gives the multivariate normal distribution; by varying λ we get distributions with different tail behaviors. In our simulations, we use MGL(λ) with $\lambda = 1, 1.5, 2$, and 5. (We take $\mu = 0$ and $\Sigma = I$.) We also do simulations using a uniform distribution inside the ball $\{y \in \mathbb{R}^p : |y| \leqslant 1\}$, which arises as the limit of MGL(λ) as $\lambda \to \infty$.

For each of these distributions, we carried out simulations for dimensions $p = 3$ and $p = 4$ and sample sizes $n = 100$ and $n = 200$. In each simulation, we generated 50,000 data sets and computed the statistics X^2, MPQ, Schott, BH, Skew, and Kurt for each data set. Tables 1–5 record, for each statistic, the proportion of values which exceed the nominal .01, .05., and .10 upper critical points which are obtained for MPQ, Schott, Skew, and Kurt from their known limiting distributions; for BH from the lognormal approximation $q_{\beta,d}(\alpha)$ of Henze and Wagner [10]; and for X^2 from the limiting distribution under normality given in (2.5).

For the simulations of this section, when computing X^2 we take the sectors to be the $g = 2^p$ orthants in $\mathbb{R}^p$, and use $c = 3$ shells in dimension $p = 3$, and $c = 2$ for $p = 4$. These values of c were chosen so that the average cell counts $n/(gc)$ are not too small. (Simulations using different types of sectors and different choices for p and c led to similar results.) For these sectors, the eigenvalues λ_i in (2.5) are given in Section 8.1. Since it is easy to generate random variables from the limit distribution in (2.5), in constructing our tables we used approximate critical points obtained from the sample quantiles of a sample of size 1,000,000. Critical values can also be obtained by numerically inverting the characteristic function as in Imhof [13].

When computing MPQ, we set $\varepsilon = .05$ in (3.1), and used the spherical harmonics $\mathcal{J}_{jl}$ in (3.2) of degrees $j = 3$ to $l = 5$. In the next section, we do several simulations in dimension $p = 2$ using degrees $j = 3$ to $l = 7$.

Table 1 shows that, when sampling from a normal distribution, the limiting distribution (2.5) is an excellent approximation to the distribution of X^2. For MGL(5) and MGL(1.5), which

Table 2
Simulation results for MGL(5)

n	Method	$p = 3$			$p = 4$		
		$\alpha = .01$	$\alpha = .05$	$\alpha = .1$	$\alpha = .01$	$\alpha = .05$	$\alpha = .1$
100	X^2	.01294	.05670	.11578	.01146	.05398	.10966
	MPQ	.00920	.04850	.09842	.00984	.04990	.10074
	Schott	.00794	.04428	.09130	.00810	.04556	.09638
	BH	.25802	.57482	.72488	.26176	.55542	.70258
	Skew	.00000	.00010	.00032	.00000	.00002	.00030
	Kurt	.68870	.97196	.99498	.84242	.99026	.99832
200	X^2	.01264	.06234	.11794	.01220	.05648	.11192
	MPQ	.00924	.04956	.09984	.00952	.04876	.09944
	Schott	.00878	.04574	.09388	.00974	.04704	.09812
	BH	.87944	.97496	.99072	.89250	.97264	.98812
	Skew	.00000	.00008	.00044	.00000	.00004	.00012
	Kurt	.99948	1.0000	1.0000	.99994	1.0000	1.0000

Table 3
Simulation results for MGL(1.5)

n	Method	$p = 3$			$p = 4$		
		$\alpha = .01$	$\alpha = .05$	$\alpha = .1$	$\alpha = .01$	$\alpha = .05$	$\alpha = .1$
100	X^2	.01232	.05102	.10470	.01068	.05100	.10460
	MPQ	.00954	.04950	.10076	.00974	.04942	.09942
	Schott	.00638	.03984	.08956	.00680	.04080	.09076
	BH	.08096	.20990	.30790	.08296	.19960	.29198
	Skew	.08760	.19696	.28350	.10086	.21794	.30800
	Kurt	.16632	.28196	.36244	.15902	.27860	.36168
200	X^2	.01198	.05612	.10774	.01210	.05492	.11212
	MPQ	.00874	.04916	.09996	.00976	.04954	.10116
	Schott	.00680	.04022	.08984	.00794	.04248	.09178
	BH	.18552	.37710	.49530	.19538	.36716	.47830
	Skew	.11030	.23942	.33940	.12824	.26908	.37582
	Kurt	.39170	.56302	.65300	.40206	.58012	.67186

represent fairly substantial deviations from normality, Tables 2 and 3 show that critical values from (2.5) result in a slightly liberal, but still reasonably accurate test. For MGL(1) and the uniform distribution on the ball, which are extreme deviations from normality, Tables 4 and 5 show that we end up with a rather liberal test.

Simulations like the above suggest that (2.5) supplies a reasonably good approximation to the distribution of X^2 for elliptically symmetric distributions which are not too far removed from the multivariate normal distribution. For fairly extreme departures from normality, we recommend using critical values obtained from the bootstrapping procedure described in Section 2. To examine the performance of this approach, we carried out the following simulation study. For samples sizes

Table 4
Simulation results for MGL(1)

n	Method	$p=3$			$p=4$		
		$\alpha=.01$	$\alpha=.05$	$\alpha=.1$	$\alpha=.01$	$\alpha=.05$	$\alpha=.1$
100	X^2	.01810	.07224	.13948	.01588	.07382	.14148
	MPQ	.01002	.05110	.10218	.01010	.05016	.10328
	Schott	.00584	.03992	.09096	.00724	.03994	.08998
	BH	.76442	.88428	.92686	.77222	.88116	.92176
	Skew	.47522	.63724	.71890	.54020	.69972	.77894
	Kurt	.84324	.92044	.94760	.85784	.93112	.95588
200	X^2	.01796	.07740	.14242	.01870	.07540	.14630
	MPQ	.01054	.05056	.10054	.01100	.05162	.10246
	Schott	.00666	.04044	.08994	.00778	.04104	.09136
	BH	.98524	.99642	.99830	.98720	.99586	.99810
	Skew	.59270	.74010	.80896	.68138	.81674	.87420
	Kurt	.99414	.99818	.99910	.99590	.99918	.99958

Table 5
Simulation results for a uniform distribution on unit ball

n	Method	$p=3$			$p=4$		
		$\alpha=.01$	$\alpha=.05$	$\alpha=.1$	$\alpha=.01$	$\alpha=.05$	$\alpha=.1$
100	X^2	.04058	.12760	.21950	.02334	.09276	.16704
	MPQ	.00878	.04806	.09830	.00962	.04942	.10098
	Schott	.01038	.05174	.10130	.01068	.05302	.10398
	BH	.92526	.98870	.99624	.94802	.99174	.99728
	Skew	.00000	.00000	.00002	.00000	.00000	.00000
	Kurt	.99996	1.0000	1.0000	1.0000	1.0000	1.0000
200	X^2	.04206	.13824	.22512	.02564	.09428	.16992
	MPQ	.00976	.05128	.09922	.00970	.05012	.10146
	Schott	.01068	.05128	.10200	.01122	.05248	.10530
	BH	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Skew	.00000	.00000	.00000	.00000	.00000	.00000
	Kurt	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$n=100$ and 200, and for dimensions $p=3$ and 4, we generated 10,000 data sets from both the MGL(1) distribution and the uniform distribution on the ball. For each data set we computed X^2 and compared this with critical values obtained from a bootstrap distribution based on 2000 bootstrap replicates. Table 6 displays the proportion of times this procedure led to rejection at levels $\alpha=.01$, .05, and .1. We see that the observed significance levels are quite close to their nominal values.

In Tables 1–5 we see that the null distribution of MPQ is close to its limiting distribution in all of our simulations. However, Schott tends to be somewhat conservative for the MGL distributions when λ is 2, 1.5, or 1.0. In Table 1, we see that Mardia's kurtosis test is rather conservative

Table 6
Observed significance levels of the bootstrap-based test for the MGL(1) distribution (#1) and the uniform distribution on the ball (#2)

n	p	c	$\alpha = .01$		$\alpha = .05$		$\alpha = .1$	
			#1	#2	#1	#2	#1	#2
100	3	3	0.0098	0.0147	0.0472	0.0628	0.0959	0.1147
100	4	2	0.0106	0.0112	0.0500	0.0494	0.0963	0.0942
200	3	3	0.0103	0.0120	0.0508	0.0516	0.0985	0.0992
200	4	2	0.0114	0.0116	0.0550	0.0478	0.1055	0.0956

for the sample sizes we study. Tables 2–5 show that the kurtosis test has considerable power to detect the departure of MGL(λ) from normality for $\lambda \neq 2$ whereas Mardia's skewness test has little or no power (but this is to be expected since elliptically symmetric distributions have zero skewness). The behavior of Mardia's skewness test in our simulations can be understood using the results of Baringhaus and Henze [2]; see in particular their Theorem 2.2 (which gives the limiting distribution under elliptical symmetry) and Remark 4.2.

5. The power of X^2 under some alternatives

In this section, we present simulation results comparing the power of X^2 with MPQ, Schott, BH, Skew, and Kurt under some alternatives. All the alternatives we consider deviate substantially from elliptical symmetry so that we might hope to be able to detect this deviation with modest sample sizes like $n = 100$ or 200. The design of the simulation study is like that in the previous section except in two respects. First, each of the empirical power values we report is based on only 500 samples from the alternative distribution; this suffices to get a reasonable idea of the relative power of the procedures. Secondly, because X^2 is not affine invariant, every data set is rotated by a random orthogonal matrix before computing X^2. That is, if Γ is a random $p \times p$ orthogonal matrix and Y denotes an $n \times p$ data matrix generated from one of the alternative distributions, we compute X^2 from $Y\Gamma$ instead of Y. This procedure eliminates the possibility that the particular orientation of the alternative distribution gives an advantage to X^2.

We have compared the power of X^2, MPQ, and Schott under a large number of alternatives. As one might expect, none of the tests is uniformly superior to the others, and both MPQ and Schott perform very well for many alternatives. Our primary purpose in this section is simply to demonstrate that the statistic X^2 would be a useful addition to the list of available procedures for testing elliptical symmetry. To achieve this purpose we will present simulation results for three alternatives where X^2 has good power relative to MPQ and Schott; given the discussion in Section 3, it is not difficult to devise such alternatives.

We begin with an interesting two-dimensional 'spiral' distribution with density, parameterized by b, given by

$$f(x_1, x_2) = \frac{1}{2\pi}\{1 + \cos[2(\theta - b \cdot r)]\}\exp(-r^2/2), \tag{5.1}$$

where r and θ are the radius and angle of (x_1, x_2) in polar coordinates. A plot of this density shows two symmetric spiral arms resembling a spiral galaxy, with the parameter b controlling

Table 7
Power against the alternative (5.1)

α	$n = 100$			$n = 200$		
	0.01	0.05	0.1	0.01	0.05	0.1
X^2	.344	.596	.712	.886	.980	.992
MPQ	.014	.036	.096	.006	.040	.086
Schott	.006	.054	.108	.008	.054	.142
BH	.028	.114	.180	.024	.166	.282
Skew	.006	.046	.090	.004	.046	.090
Kurt	.000	.014	.042	.010	.028	.064

how tightly these arms wrap around the center. As b increases, the moments of this distribution converge rapidly to those of the bivariate normal distribution $N_2(0, I)$. In our simulations we take $b = 2$ which produces a distribution whose moments up to order four are close to those of the normal distribution, but whose contours are very far from being elliptical. In computing X^2, we use $g = 4$ sectors (the quadrants) and $c = 5$.

From the results in Table 7, we see that our method has great power to detect the spiral pattern of this distribution; BH is the only other test that has some power here. Since Schott, Skew and Kurt are designed to detect specific departures of the moments from those expected under elliptical symmetry or normality, respectively, their poor performance here is easily understood (since the spiral distribution has moments very close to those of the normal distribution). The reason for MPQ's lack of power is a little more subtle. In dimension $p = 2$, MPQ reduces to a test that the direction vectors $u_i = z_i/|z_i|$ are uniformly distributed on the circle. For the spiral distribution, the distribution of u_i differs somewhat from uniform, but the departure is mainly in the direction of a second degree harmonic for which MPQ has no power. (See the discussion in Section 3.) We note that BH has power approaching 1 as $n \to \infty$ for all alternatives, but for the spiral distribution requires $n > 1000$ to have power comparable to X^2 with $n = 200$.

Our second alternative is a three-dimensional distribution with density

$$f(x_1, x_2, x_3) = \frac{2}{(2\pi)^{3/2}} \exp(-r^2/2) I(x_1 x_2 x_3 (r^2 - m) < 0), \tag{5.2}$$

where $r^2 = x_1^2 + x_2^2 + x_3^2$ and m is the median of the χ_3^2 distribution. This density is obtained by multiplying the $N_3(0, I)$ density by an indicator function which, in each orthant, wipes out (sets to zero) the density either inside or outside of the sphere of radius $\sqrt{m}$ about the origin. The simulation results for this alternative are given in Table 8. We see that Skew has excellent power. Among the tests of elliptical symmetry, X^2 using $g = 8$ sectors (the orthants) and $c = 3$ shells has moderate power, but MPQ and Schott have no power at all. We can increase the power of X^2 in this situation by using more sectors. Table 8 gives the power for $g = 12$ (20) and $c = 2$ (3). When g is 12 (20), the sectors correspond to the faces of a dodecahedron (icosahedron). For $g = 8$ we used critical values from the limiting distribution, but for $g = 12$ and 20, we used the bootstrap testing procedure with 2000 bootstrap replicates. We did this because for this choice of sectors we do not know the exact eigenvalues λ_i needed in (2.5), and also because, even if we did, for many of the situations in Table 8 the average cell counts are too small to justify use of the limiting distribution.

Table 8
Power against the alternative (5.2)

α	$n = 100$			$n = 200$		
	0.01	0.05	0.1	0.01	0.05	0.1
$X^2(g = 8, c = 3)$	0.152	0.302	0.429	0.296	0.488	0.591
$X^2(g = 12, c = 2)$	0.450	0.664	0.742	0.784	0.870	0.916
$X^2(g = 12, c = 3)$	0.264	0.484	0.588	0.636	0.784	0.834
$X^2(g = 20, c = 2)$	0.534	0.758	0.848	0.952	0.982	0.990
$X^2(g = 20, c = 3)$	0.288	0.556	0.712	0.814	0.930	0.958
MPQ	0.016	0.066	0.126	0.024	0.079	0.162
Schott	0.005	0.020	0.060	0.004	0.032	0.085
BH	0.075	0.240	0.393	0.286	0.638	0.808
Skew	0.386	0.703	0.828	0.973	0.999	1.000
Kurt	0.009	0.029	0.064	0.008	0.040	0.077

Table 9
Power against the alternative (5.3)

α	$n = 100$			$n = 200$		
	.01	.05	.1	.01	.05	.1
$X^2(g = 4, c = 5)$	.280	.444	.538	.416	.564	.618
$X^2(g = 5, c = 4)$	.372	.634	.752	.722	.904	.950
$X^2(g = 7, c = 3)$	.574	.814	.868	.902	.976	.984
$X^2(g = 8, c = 3)$	.612	.804	.882	.94	.982	.994
$X^2(g = 12, c = 2)$	.592	.776	.842	.928	.980	.992
MPQ	.012	.064	.110	.018	.072	.150
Schott	.330	.584	.726	.768	.910	.960
BH	1.00	1.00	1.00	1.00	1.00	1.00
Skew	.000	.000	.000	.000	.000	.000
Kurt	1.00	1.00	1.00	1.00	1.00	1.00

Our third alternative distribution has density

$$f(x_1, x_2) = \begin{cases} 1/\pi & \text{if } r \in [0, 1) \text{ and } \theta \in [0, \pi/2) \cup [\pi, 3\pi/2) \\ & \text{or } r \in [1, \sqrt{2}) \text{ and } \theta \in [\pi/2, \pi) \cup [3\pi/2, 2\pi), \\ 0 & \text{otherwise,} \end{cases} \tag{5.3}$$

where r and θ are the radius and the angle of (x_1, x_2). Simulation results for this distribution are given in Table 9. For this distribution, the lengths $|z_i|$ have a very different distribution from that obtained under normality, so for X^2 we used bootstrap critical values based on 2000 replicates. The distribution (5.3) has zero skewness and the angle θ is almost uniform on $[0, 2\pi]$, so it is not surprising that Skew and MPQ have little power here. But Schott, BH, and Kurt have very high power against this alternative. The power of X^2 varies considerably depending on the choice of g and c, but for a range of values we get power comparable to or exceeding that of Schott.

We shall present two more alternatives. Our goal in presenting these last alternatives is different from that in the earlier ones. For these alternatives, our X^2 test does not do well for the sample

Table 10
Power against alternative (5.4)

n	Method	$p=3$			$p=4$		
		$\alpha=.01$	$\alpha=.05$	$\alpha=.1$	$\alpha=.01$	$\alpha=.05$	$\alpha=.1$
100	X^2	0.016	0.062	0.136	0.018	0.086	0.148
	MPQ	0.208	0.394	0.520	0.244	0.420	0.546
	Schott	0.318	0.598	0.726	0.444	0.726	0.848
	BH	0.052	0.140	0.260	0.066	0.202	0.318
	Skew	0.050	0.120	0.172	0.070	0.140	0.218
	Kurt	0.054	0.122	0.154	0.058	0.136	0.190
200	X^2	0.010	0.060	0.126	0.028	0.096	0.188
	MPQ	0.560	0.762	0.850	0.684	0.868	0.932
	Schott	0.842	0.958	0.978	0.920	0.972	0.986
	BH	0.116	0.344	0.512	0.200	0.462	0.604
	Skew	0.076	0.148	0.220	0.090	0.200	0.282
	Kurt	0.084	0.178	0.256	0.164	0.270	0.342

Table 11
Power against the alternative (5.5)

n	Method	α		
		0.01	0.05	0.1
100	X^2(orthants)	.028	.120	.198
	X^2(product)	.170	.352	.458
	MPQ	.844	.964	.988
	Schott	.302	.602	.750
	BH	.076	.270	.408
	Skew	.008	.054	.092
	Kurt	.004	.024	.064
200	X^2(orthants)	.082	.186	.306
	X^2(product)	.528	.748	.830
	MPQ	1.00	1.00	1.00
	Schott	.834	.966	.990
	BH	.412	.760	.910
	Skew	.018	.058	.100
	Kurt	.002	.040	.102

sizes $n \leqslant 200$ considered in this paper, but MPQ and Schott have excellent power which greatly exceeds that of the multivariate normality tests BH, Skew, and Kurt; see the simulation results in Tables 10 and 11. (We note that for larger values of n, we can increase the power of X^2 by using more cells.) These results and the earlier results for the alternative (5.1) in Table 7 where X^2 outperformed the normality tests suggest that X^2, MPQ, and Schott will also be useful as specialized tests of multivariate normality, specialized in the sense that they have power against alternatives which violate elliptical symmetry.

Our next alternative is a multivariate ℓ_1-norm symmetric distribution which is a modification of a distribution introduced in Chapter 5 of Fang et al. [8]. For this distribution, the data vectors have the form in (2.1), but with U uniformly distributed on the surface of the unit ℓ_1-ball in $\mathbb{R}^p$. In our simulations, we take $\xi^2 \sim \chi_p^2$ which gives a distribution with ℓ_1-norm contours whose tails decay at the same rate as the multivariate normal distribution. In summary, under this alternative, the data y satisfies

$$y \stackrel{\text{d}}{=} \xi U \quad \text{where } \xi \text{ and } U \text{ are independent, } \xi^2 \sim \chi_p^2 \text{ and}$$

$$U \text{ is uniformly distributed on } \left\{ y \in \mathbb{R}^p : \sum\nolimits_{i=1}^{p} |y_i| = 1 \right\}. \tag{5.4}$$

It is straightforward to generate random vectors from this distribution. Simulation results for this alternative in dimensions $p = 3$ and 4 are reported in Table 10.

Our last alternative has density

$$f(x_1, x_2, x_3, x_4) = \frac{1}{2\pi^2} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2 + x_4^2) \right) I(x_1 x_2 x_3 x_4 > 0) \tag{5.5}$$

which is the $N_4(0, I)$ density restricted to the orthants where $x_1 x_2 x_3 x_4 > 0$. The simulation results for this alternative are given in Table 11. We present results for two versions of our chi-square statistic: X^2(orthants) uses the $g = 16$ orthants as sectors (with $c = 2$), and X^2(product) uses $g = 49$ sectors (with $c = 1$) obtained using the product construction in Section 9 with $g_1 = g_2 = 7$. In this situation, the power of X^2(product) is much better than that of X^2(orthants), but still falls well short of the power of MPQ and Schott.

6. Proofs

In this section we do two things. First, we prove the ancillarity result for the Gram–Schmidt transformation stated in Section 2. Secondly, we derive the limiting distribution of X^2 stated earlier in (2.5). (This second item is the main business of the section.)

First, some brief remarks on notation. We will use I, e, and 0 to denote an identity matrix, a column vector of ones, and a column vector or matrix of zeros, respectively. The dimensions will usually be clear from context, but will be specified by subscripts if necessary. Unless otherwise noted vectors will be column vectors, but for convenience they will be written in text as row vectors.

Let the $n \times p$ data matrix $Y = (y_{ij})$ denote the original (raw) data matrix whose n rows $y_1, y_2, \ldots, y_n$ are a random sample from a p-variate elliptically symmetric distribution as in (2.2). Let Z be the $n \times p$ matrix whose rows are the scaled residuals $z_1, z_2, \ldots, z_n$ from (1.1).

Lemma 6.1. *If the scaled residuals Z are obtained using the Gram–Schmidt transformation, then the distribution of Z does not depend on μ or $\Sigma = AA^t$.*

Proof. Since the distribution of $y_i - \bar{y}$ does not depend on μ, it is clear that the distribution of Z also does not, which proves part of the lemma. Thus, we can take $\mu = 0$ in (2.2). Let $Z(A)$ be the matrix of scaled residuals obtained from $y_i = \xi_i A U_i$, $i = 1, \ldots, n$, where ξ_i and U_i are iid copies of ξ and U in (2.2). Let $\mathcal{I}$ be the set of $p \times p$ invertible matrices; $\mathcal{O}$, the $p \times p$ orthogonal matrices; and $\mathcal{L}$, the lower triangular matrices with positive diagonal elements. For $A_1, A_2 \in \mathcal{I}$, we define $A_1 \equiv A_2$ to mean $Z(A_1) \stackrel{\text{d}}{=} Z(A_2)$. Since the distribution of U is spherically symmetric, that is, $U \stackrel{\text{d}}{=} \Gamma U$ for $\Gamma \in \mathcal{O}$, it follows that $A \equiv A\Gamma$ for $A \in \mathcal{I}$ and $\Gamma \in \mathcal{O}$. In (1.1), the Gram–Schmidt transformation takes $R(S) = L^{-1}$ where L is the unique matrix in $\mathcal{L}$ satisfying $S = LL^t$

(the Choleski decomposition). From this, it is easy to verify that

$$R(MSM^t) = R(S)M^{-1} \quad \text{for positive definite } S \text{ and } M \in \mathcal{L}. \tag{6.1}$$

Since a linear transformation of the data $y \mapsto My$ transforms the sample covariance matrix according to the "sandwich" rule $S \mapsto MSM^t$, the fact (6.1) implies $Z(MA) = Z(A)$ for $A \in \mathcal{I}$ and $M \in \mathcal{L}$ so that $MA \equiv A$. For any $A \in \mathcal{I}$, there exists $M \in \mathcal{L}$ and $\Gamma \in \mathcal{O}$ such that $A = M\Gamma$. Thus $I \equiv I\,\Gamma = \Gamma \equiv M\Gamma = A$ and the proof is complete. □

Proof of (2.5). From this point on we assume normality: $y_1, \ldots, y_n$ are iid $N(\mu, \Sigma)$. We will continue to assume use of the Gram–Schmidt transformation in (1.1), but in fact, according to Lemma 3.1 in Huffer and Park [12], under normality the distribution of Z is the same for all choices of the function $R(S)$.

In Section 2 we defined the cells used in the construction of our chi-square statistic. For the purposes of this proof, we require a somewhat more general notation for the cells. Let $\theta = (\mu, \Sigma)$ denote the parameters of the multivariate normal distribution. Let $q = (q_1, q_2, \ldots, q_{c-1})$ where $0 < q_1 < q_2 < \cdots < q_{c-1}$. Let $z = z(y, \theta) = R(\Sigma)(y - \mu)$. With G_i as in Section 2, define $G_i(\theta) = \{y : z \in G_i\}$ for $1 \leqslant i \leqslant g$. Define $S_j(\theta, q) = \{y : q_{j-1} < |z|^2 \leqslant q_j\}$ for $1 \leqslant j \leqslant c$ where we take $q_0 = 0$ and $q_c = \infty$. Note that $|z|^2 = (y - \mu)^t \Sigma^{-1} (y - \mu)$. For $\pi = (i, j)$ with $1 \leqslant i \leqslant g$ and $1 \leqslant j \leqslant c$, we define

$$\Lambda_\pi(\theta, q) = G_i(\theta) \cap S_j(\theta, q).$$

Given data $y_1, y_2, \ldots, y_n$, we define $\theta_n = (\bar{y}, S)$ and $z_i = z(y_i, \theta_n)$. Note that $z_1, \ldots, z_n$ defined in this way are exactly the scaled residuals in (1.1). Let $q_n = (q_{n1}, q_{n2}, \ldots, q_{n\,c-1})$ where q_{nj} are as in (2.3), and define the vector of cells $\Lambda_n = (\Lambda_{n\pi})$ with $\Lambda_{n\pi} = \Lambda_\pi(\theta_n, q_n)$. In this $gc \times 1$ vector, the cells, indexed by $\pi = (i, j)$ with $1 \leqslant i \leqslant g$ and $1 \leqslant j \leqslant c$, are grouped by the value of j and then ordered by i within each group. In other words, cell π occupies position $i + g(j - 1)$. Define the vector of cell counts $U_n = (U_{n\pi})$ by

$$U_{n\pi} = \sum_{k=1}^{n} I(y_k \in \Lambda_{n\pi})$$

and let X^2 be as in (2.4). Note that the cell counts $U_{n\pi}$ are the same as those defined in (2.3), but for convenience in the proof we have changed the definition of the cells $\Lambda_{n\pi}$ to be in terms of the original values y_k instead of the scaled residuals z_k.

Since X^2 can be expressed as a function of Z whose distribution does not depend on $\theta = (\mu, \Sigma)$, we can assume without loss of generality that $\theta = \theta_0 = (0, I)$. We assume from this point on that $y_1, y_2, \ldots, y_n$ are iid $N(0, I)$.

For an arbitrary measurable region $\Upsilon \subset \mathbb{R}^p$, we define $U_n(\Upsilon) = \sum_{i=1}^n I(y_i \in \Upsilon)$ and define $\mathrm{P}(\Upsilon, \theta)$ to be the probability assigned to Υ by the normal distribution with parameter θ. We also let $D(\Upsilon)$ be $\partial \mathrm{P}(\Upsilon, \theta)/\partial\theta$ evaluated at $\theta = \theta_0$. For a vector of regions $\Upsilon = (\Upsilon_i)$ we use the obvious vector analogs of the above definitions so that $U_n(\Upsilon) = (U_n(\Upsilon_i))$, $\mathrm{P}(\Upsilon, \theta)$ is a vector of probabilities, and $D(\Upsilon)$ is a matrix of partial derivatives. Our vector of cell counts U_n may now be written as $U_n(\Lambda_n)$. Finally, we define the process $V_n(\Upsilon) = n^{-1/2}(U_n(\Upsilon) - n\mathrm{P}(\Upsilon, \theta_0))$.

Let $q_0 = (q_{01}, q_{02}, \ldots, q_{0\,c-1})$ be the vector of population quantiles of the χ^2_p distribution which divide this distribution into c intervals of equal probability $1/c$. Define the vector of limiting cells

$\Lambda_0 = (\Lambda_{0\pi})$ with $\Lambda_{0\pi} = \Lambda_\pi(\theta_0, q_0)$. Clearly $\theta_n \xrightarrow{p} \theta_0$ and $q_n \xrightarrow{p} q_0$ so that $\Lambda_n \longrightarrow \Lambda_0$ in the sense that $P(\Lambda_{n\pi} \triangle \Lambda_{0\pi}, \theta_0) \xrightarrow{p} 0$ for all cells π.

With these definitions we may state the following approximation lemma. The proof (which we omit) is essentially identical to that of Lemma 3.2 in Huffer and Park [12] and relies on basic ideas from the theory of empirical processes (see [21]). □

Lemma 6.2.
$n^{-1/2}\,(U_n(\Lambda_n) - n\mathrm{P}(\Lambda_n, \theta_n)) = V_n(\Lambda_0) - D(\Lambda_0)\sqrt{n}(\theta_n - \theta_0) + o_p(1)\,.$

Suppose the coordinates of $\theta = (\mu, \Sigma)$ are arranged so that $\theta = (\mu, \sigma, \rho)$, where $\sigma = (\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp})$ and $\rho = (\sigma_{12}, \sigma_{13}, \ldots, \sigma_{(p-1)p})$ are the diagonal and off-diagonal elements of Σ, respectively. Then we may write $D(\Lambda_0) = (A, B, C)$, where

$$A = \frac{\partial P(\Lambda_0, \theta)}{\partial \mu}, \quad B = \frac{\partial P(\Lambda_0, \theta)}{\partial \sigma}, \quad C = \frac{\partial P(\Lambda_0, \theta)}{\partial \rho},$$

all evaluated at $\theta = \theta_0$. For any vector $x = (x_1, x_2, \ldots, x_p)$, we define the column vectors $s(x) = (x_1^2, x_2^2, \ldots, x_p^2)$, and $r(x) = (x_1x_2, x_1x_3, \ldots, x_{p-1}x_p)$ with dimensions p and $p(p-1)/2$, respectively. Then since $\theta_n = (\bar{y}, S)$ and $S = Y^tY/n + o_p(n^{-1/2})$, we obtain from Lemma 6.2 that

$$\begin{aligned} &n^{-1/2}\,(U_n(\Lambda_n) - n\mathrm{P}(\Lambda_n, \theta_n)) \\ &\quad = V_n(\Lambda_0) - \sqrt{n}\left(A\bar{y} + B\sum_{i=1}^{n}(s(y_i) - e)/n + C\sum_{i=1}^{n} r(y_i)/n\right) + o_p(1). \end{aligned} \tag{6.2}$$

Define the projection matrix $\Pi = I - E$ where E is the $gc \times gc$ block diagonal matrix given by

$$E = \frac{1}{g}\,\mathrm{diag}(e_g e_g^t, \ldots, e_g e_g^t)\,.$$

It is easily seen that $\Pi\mathrm{P}(\Lambda_n, \theta_n) = 0$ and $\Pi U_n(\Lambda_n) = U_n(\Lambda_n) - np_0e + O(1)$ so that applying Π to both sides of (6.2) yields

$$\begin{aligned} &n^{-1/2}\,(U_n(\Lambda_n) - np_0\, e) \\ &\quad = \Pi\left\{V_n(\Lambda_0) - \sqrt{n}\left(A\bar{y} + B\sum_{i=1}^{n}(s(y_i) - e)/n + C\sum_{i=1}^{n} r(y_i)/n\right)\right\} + o_p(1). \end{aligned} \tag{6.3}$$

The central limit theorem implies that

$$\left(V_n(\Lambda_0), \sqrt{n}\bar{y}, n^{-1/2}\sum_{i=1}^{n}(s(y_i) - e), n^{-1/2}\sum_{i=1}^{n} r(y_i)\right)$$

is asymptotically normal with mean zero and covariance matrix (found after some calculation) equal to

$$\begin{pmatrix} p_0(I - p_0ee^t) & A & 2B & C \\ A^t & I & 0 & 0 \\ 2B^t & 0 & 2I & 0 \\ C^t & 0 & 0 & I \end{pmatrix}. \tag{6.4}$$

In this covariance matrix, the diagonal blocks and the off-diagonal blocks of zeros are immediate. The occurrence of the matrices A, $2B$, and C among the off-diagonal blocks comes from the following formulas:

$$\begin{aligned} \mathrm{Var}(y, I(y \in \Upsilon)) &= \int_\Upsilon x f_0(x)\,dx &&= \tfrac{\partial}{\partial \mu} P(\Upsilon, \theta_0) \\ \mathrm{Var}(s(y), I(y \in \Upsilon)) &= \int_\Upsilon s(x) f_0(x)\,dx - P(\Upsilon, \theta_0)e &&= 2\tfrac{\partial}{\partial \sigma} P(\Upsilon, \theta_0) \\ \mathrm{Var}(r(y), I(y \in \Upsilon)) &= \int_\Upsilon r(x) f_0(x)\,dx &&= \tfrac{\partial}{\partial \rho} P(\Upsilon, \theta_0) \end{aligned} \tag{6.5}$$

which are valid for any region $\Upsilon \subset \mathbb{R}^p$. Here $f_0(x)$ denotes the multivariate normal density for $\theta = \theta_0 = (0, I)$.

The right-hand sides of the formulas in (6.5) are easy consequences of the general results about matrix/vector derivatives given in Dwyer [5]; in particular see formulas (11.1), (11.3), and (11.8) of his Table 2. Formula (11.1) leads directly to the expression for $(\partial/\partial\mu)P(\Upsilon, \theta_0)$. Formulas (11.3) and (11.8) lead to

$$\frac{\partial}{\partial \Sigma} P(\Upsilon, \theta_0) = \frac{1}{2}\int_\Upsilon x x^t f_0(x)\,dx - \frac{1}{2} P(\Upsilon, \theta_0) I. \tag{6.6}$$

Here we are differentiating with respect to the symmetric matrix $\Sigma = (\sigma_{ij})$; for any function $g(\theta)$ we define $\partial g/\partial\Sigma$ to be the $p \times p$ symmetric matrix with entries

$$\left(\frac{\partial g}{\partial \Sigma}\right)_{ij} = \begin{cases} \dfrac{\partial g}{\partial \sigma_{ii}} & \text{for } i = j \\ \dfrac{1}{2}\dfrac{\partial g}{\partial \sigma_{ij}} & \text{for } i \neq j. \end{cases}$$

With this definition, examination of the diagonal and off-diagonal elements in (6.6) leads directly to the formulas for $(\partial/\partial\sigma)P(\Upsilon, \theta_0)$ and $(\partial/\partial\rho)P(\Upsilon, \theta_0)$ in (6.5).

From (6.3) and the limiting covariance in (6.4) we obtain

$$(U_n(\Lambda_n) - np_0\, e)\,/\sqrt{np_0} \xrightarrow{\mathrm{d}} N(0, \Psi),$$

where

$$\begin{aligned} \Psi &= p_0^{-1}\Pi\left(p_0(I - p_0 e e^t) - AA^t - 2BB^t - CC^t\right)\Pi \\ &= I - E - QQ^t \quad \text{where } Q = \sqrt{gc}\,\Pi\left(A, \sqrt{2}B, C\right). \end{aligned} \tag{6.7}$$

The limiting distribution of X^2 will thus be that of a linear combination of chi-squared random variables with coefficients supplied by the nonzero eigenvalues of Ψ. Let $\lambda_1, \ldots, \lambda_m$ be the nonzero eigenvalues of QQ^t. Since the columns of E are orthogonal to those of Q, the eigenvalues of Ψ are easily seen to be $1 - \lambda_i$ for $i = 1, \ldots, m$, and 1 with multiplicity $c(g-1) - m$. This leads immediately to the result in (2.5). Note that the nonzero eigenvalues of QQ^t are the same as those of Q^tQ.

7. Calculation of the eigenvalues of QQ^t

For some groups $\mathcal{G}$, we have been able to compute the nonzero eigenvalues of QQ^t and thus obtain the limiting distribution of X^2 explicitly. These results are listed in the next section. In this section we describe the approach used to compute the eigenvalues.

If $y \sim N(0, I)$, then $y \stackrel{\mathrm{d}}{=} \xi U$ where $\xi^2 \sim \chi^2_p$, U is uniformly distributed on Ω_p, and ξ and U are independent. Using this result, for $\pi = (i, j)$ we obtain

$$\begin{aligned} E y_k^r y_\ell^s I(y \in \Lambda_{0\pi}) &= E\left[\xi^{r+s} U_k^r U_\ell^s I\{\xi U \in G_i(\theta_0) \cap S_j(\theta_0, q_0)\}\right] \\ &= \left(E\xi^{r+s} I(q_{0j-1} < \xi^2 \leqslant q_{0j})\right)\left(E U_k^r U_\ell^s I(U \in G_i)\right) \\ &= \left(\frac{E\xi^{r+s} I(q_{0j-1} < \xi^2 \leqslant q_{0j})}{E\xi^{r+s}}\right)\left(E y_k^r y_\ell^s I(y \in G_i)\right). \end{aligned} \tag{7.1}$$

For $j = 1, \ldots, c$, define

$$\begin{aligned} a_j &= E\xi I(q_{0j-1} < \xi^2 \leqslant q_{0j})/E\xi = F_{p+1}(q_{0j}) - F_{p+1}(q_{0j-1}), \quad \text{and} \\ b_j &= E\xi^2 I(q_{0j-1} < \xi^2 \leqslant q_{0j})/E\xi^2 = F_{p+2}(q_{0j}) - F_{p+2}(q_{0j-1}), \end{aligned} \tag{7.2}$$

where F_k is the distribution function of the χ^2_k distribution. From (6.5), (7.1), and (7.2), we obtain (for $\pi = (i, j)$)

$$\begin{aligned} \frac{\partial}{\partial\mu} P(\Lambda_{0\pi}, \theta_0) &= a_j \int_{G_i} x f_0(x)\, dx, \\ \frac{\partial}{\partial\sigma} P(\Lambda_{0\pi}, \theta_0) &= \frac{1}{2} b_j \int_{G_i} s(x) f_0(x)\, dx - \frac{1}{2} p_0 e, \\ \frac{\partial}{\partial\rho} P(\Lambda_{0\pi}, \theta_0) &= b_j \int_{G_i} r(x) f_0(x)\, dx. \end{aligned} \tag{7.3}$$

This allows us to write A, B, and C as Kronecker products as follows. Define the $c \times 1$ vectors $a = (a_1, \ldots, a_c)$ and $b = (b_1, \ldots, b_c)$. Let $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ be matrices with g rows whose ith rows are given by $\int_{G_i} x f_0(x)\, dx$, $\int_{G_i} s(x) f_0(x)\, dx$, and $\int_{G_i} r(x) f_0(x)\, dx$, respectively. That is, we define

$$\tilde{A}_{ik} = \int_{G_i} x_k f_0(x)\, dx, \quad \tilde{B}_{ik} = \int_{G_i} x_k^2 f_0(x)\, dx, \quad \tilde{C}_{ik} = \int_{G_i} x_\ell x_m f_0(x)\, dx, \tag{7.4}$$

where in the last integral we take k to be the position of the term $x_\ell x_m$ in the list $r(x)$. With this notation, we can rewrite (7.3) as

$$A = a \otimes \tilde{A}, \quad B = \tfrac{1}{2} b \otimes \tilde{B} - \tfrac{1}{2} p_0, \quad C = b \otimes \tilde{C}. \tag{7.5}$$

Here we introduce the convention that, for any matrix M and scalar d, the matrix $M - d$ is that obtained by subtracting d from all the entries of M.

The projection matrix Π can also be written as a Kronecker product: $\Pi = I_c \otimes \tilde{\Pi}$ where $\tilde{\Pi} = I_g - g^{-1} e_g e_g^t$. Since $f_0(x)$ has mean 0 and covariance matrix I, we know

$$\sum_i \int_{G_i} x f_0(x)\, dx = 0, \quad \sum_i \int_{G_i} s(x) f_0(x)\, dx = e, \quad \sum_i \int_{G_i} r(x) f_0(x)\, dx = 0,$$

which implies $\tilde{\Pi}\tilde{A} = \tilde{A}$, $\tilde{\Pi}\tilde{B} = \tilde{B} - g^{-1}$, and $\tilde{\Pi}\tilde{C} = \tilde{C}$. Thus, substituting (7.5) in (6.7) leads to

$$\begin{aligned} Q &= \sqrt{gc}\,(a \otimes \tilde{A},\ 2^{-1/2} b \otimes (\tilde{B} - g^{-1}),\ b \otimes \tilde{C}) \\ &= \sqrt{gc}\,(a \otimes \tilde{A},\ b \otimes \tilde{D}) \quad \text{where } \tilde{D} = \left(2^{-1/2}(\tilde{B} - g^{-1}), \tilde{C}\right). \end{aligned} \tag{7.6}$$

Therefore

$$Q^t Q = g \begin{pmatrix} S_{aa}\tilde{A}^t\tilde{A} & S_{ab}\tilde{A}^t\tilde{D} \\ S_{ab}\tilde{D}^t\tilde{A} & S_{bb}\tilde{D}^t\tilde{D} \end{pmatrix} \quad \text{where } S_{aa} = ca^t a,\ S_{bb} = cb^t b, \ \text{ and } \ S_{ab} = ca^t b\,. \tag{7.7}$$

(Note that $S_{aa} = S_{bb} = S_{ab} = 1$ when $c = 1$.) If $\tilde{A}^t\tilde{D} = 0$, then the matrix $Q^t Q$ is block diagonal and the desired eigenvalues have a simple form. Let $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}_A$, and $\boldsymbol{\lambda}_D$ be row vectors consisting of the nonzero eigenvalues of $Q^t Q$, $g\tilde{A}^t\tilde{A}$, and $g\tilde{D}^t\tilde{D}$, respectively. Then

$$\boldsymbol{\lambda} = (S_{aa}\boldsymbol{\lambda}_A, S_{bb}\boldsymbol{\lambda}_D)\,. \tag{7.8}$$

The condition $\tilde{A}^t\tilde{D} = 0$ turns out to be satisfied for most of the groups we consider in the next section.

The matrices $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ are easily computed using the following observation. Let Γ be an arbitrary orthogonal matrix, and Υ an arbitrary region in $\mathbb{R}^p$. Since $f_0(\Gamma x) = f_0(x)$ and the determinant of Γ equals one, it is clear that

$$\int_{\Gamma\Upsilon} x f_0(x)\,dx = \Gamma \int_{\Upsilon} x f_0(x)\,dx, \quad \text{and}$$
$$\int_{\Gamma\Upsilon} xx^t f_0(x)\,dx = \Gamma \left(\int_{\Upsilon} xx^t f_0(x)\,dx \right) \Gamma^t\,.$$

Define

$$v = \int_G x f_0(x)\,dx \quad \text{and} \quad M = \int_G xx^t f_0(x)\,dx\,. \tag{7.9}$$

Since $G_i = \Gamma_i G$, (7.4) and the above facts imply

$$\tilde{A}_{ik} = (\Gamma_i v)_k, \quad \tilde{B}_{ik} = (\Gamma_i M \Gamma_i^t)_{kk}, \quad \tilde{C}_{ik} = (\Gamma_i M \Gamma_i^t)_{\ell m}, \tag{7.10}$$

where again k in the last integral denotes the position of $x_\ell x_m$ in $r(x)$. Thus, computation of $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ essentially boils down to the computation of v and M. For some groups and fundamental regions, v and M can be given in closed form. In other cases they must be approximated by numerical integration or Monte Carlo.

In summary, here is the procedure for computing the eigenvalues λ_i required in (2.5): first compute v and M in (7.9). Use these to compute $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ in (7.10). Then form the matrix $\tilde{D}$ in (7.6). If $\tilde{A}^t\tilde{D} = 0$, then the eigenvalues are given by (7.8). If not, the required eigenvalues are those of the matrix in (7.7). We often use the symbolic math package Maple to assist us in carrying out this process.

8. Examples. The eigenvalues for various groups

8.1. *The case of orthants*

We now consider the situation where the group $\mathcal{G}$ consists of the $g = 2^p$ possible diagonal matrices whose diagonal entries are ± 1, and the fundamental region G is the positive orthant

$$G = \{x \in \mathbb{R}^p : x_i > 0 \text{ for } i = 1, \ldots, p\}.$$

Then $G_1, \ldots, G_g$ are just the orthants. In this case, the eigenvalue calculations can be carried out completely ‘by hand’ (without resorting to Maple), so we shall give them in detail. (In fact, in this case the eigenvalues can be obtained without even introducing the group $\mathcal{G}$, but we shall use it to illustrate the previous discussion.)

The vector v and matrix M are easily calculated:

$$v_i = \int_G x_i f_0(x)\, dx = \frac{1}{2^{p-1}} \int_0^\infty z\phi(z)\, dz = \frac{1}{2^p}\sqrt{\frac{2}{\pi}},$$
$$M_{ii} = \int_G x_i^2 f_0(x)\, dx = \frac{1}{2^{p-1}} \int_0^\infty z^2\phi(z)\, dz = \frac{1}{2^p},$$
$$M_{ij} = \int_G x_i x_j f_0(x)\, dx = \frac{1}{2^{p-2}} \left(\int_0^\infty z\phi(z)\, dz\right)^2 = \frac{1}{2^p}\cdot\frac{2}{\pi} \quad \text{for } i \neq j,$$

where ϕ is the standard normal density. Now, using (7.10) and keeping in mind the special form of the matrices Γ_i, we obtain

$$\tilde{A}_{ik} = \frac{1}{2^p}\sqrt{\frac{2}{\pi}}(\Gamma_i)_{kk}, \quad \tilde{B}_{ik} = \frac{1}{2^p}, \quad \tilde{C}_{ik} = \frac{1}{2^{p-1}\pi}(\Gamma_i)_{\ell\ell}(\Gamma_i)_{mm}. \tag{8.1}$$

Thus $\tilde{B} - g^{-1} = 0$ so that $\tilde{D} = (0, \tilde{C})$. Using (8.1), it is now easy to see that $\tilde{A}^t\tilde{C} = 0$ (so that (7.8) applies) and that

$$g\tilde{A}^t\tilde{A} = \frac{2}{\pi} I_p \quad \text{and} \quad g\tilde{C}^t\tilde{C} = \frac{4}{\pi^2} I_{p(p-1)/2}$$

whose eigenvalues are $2/\pi$ with multiplicity p, and $4/\pi^2$ with multiplicity $p(p-1)/2$, respectively. Thus, using (7.8), we conclude that there are $m = p(p+1)/2$ nonzero eigenvalues λ_i in (2.5) which are $(2/\pi)S_{aa}$ with multiplicity p, and $(4/\pi^2)S_{bb}$ with multiplicity $p(p-1)/2$.

Define

$$a^* = \frac{2c}{\pi}\sum_{j=1}^{c} a_j^2 \quad \text{and} \quad b^* = \frac{4c}{\pi^2}\sum_{j=1}^{c} b_j^2 .$$

We have shown in this case that

$$X^2 \xrightarrow{\text{d}} W_0 + (1-a^*)W_1 + (1-b^*)W_2 \quad \text{as } n \to \infty,$$

where W_0, W_1, W_2 are independent chi-squared variates with $c(2^p-1) - p(p+1)/2$, p, and $p(p-1)/2$ degrees of freedom, respectively.

8.2. The bivariate case

Suppose $p = 2$. Let the group $\mathcal{G}$ be the cyclic group of order g consisting of rotations by angles which are multiples of $2\pi/g$. For this group, we may take the fundamental region to be the set of points whose angle in polar coordinates is between 0 and $2\pi/g$, that is,

$$G = \{(r\ \cos(\theta), r\ \sin(\theta)) : 0 < \theta < 2\pi/g \text{ and } r > 0\}.$$

With this G, the quantities v and M in (7.9) are easily calculated, and we may follow the recipe described earlier (with an assist from Maple) to extract the nonzero eigenvalues of Q^tQ. For $g \neq 3$ we have $\tilde{A}^t\tilde{D} = 0$ so that the eigenvalues have the form in (7.8). Our results are reported below.

For $g \geqslant 5$, the limiting distribution has the form in (2.5) with $m = 4$ nonzero eigenvalues given by

$$\lambda_1 = \lambda_2 = \frac{g^2 S_{aa}}{8\pi}\left(1 - \cos\left(\frac{2\pi}{g}\right)\right),$$
$$\lambda_3 = \lambda_4 = \frac{g^2 S_{bb}}{16\pi^2}\left(1 - \cos\left(\frac{4\pi}{g}\right)\right). \tag{8.2}$$

When g is 2, 3, or 4, the eigenvalues do not follow this pattern, and these cases must be dealt with separately. When $g = 2$, we have $m = 1$, and the single nonzero eigenvalue is

$$\lambda_1 = \frac{2S_{aa}}{\pi}.$$

When $g = 4$, the regions G_i are simply the four quadrants, so that this is a special case of orthants considered earlier. Setting $p = 2$ in the earlier results we find that $m = 3$, and the eigenvalues are

$$\lambda_1 = \lambda_2 = \frac{2S_{aa}}{\pi}, \quad \lambda_3 = \frac{4S_{bb}}{\pi^2}.$$

The case $g = 3$ is messier. In this one case we must extract the eigenvalues of the matrix in (7.7). When $c \geqslant 2$, there are $m = 4$ nonzero eigenvalues given by

$$\lambda_1 = \lambda_2 = \frac{27}{64\pi^2}\left(\psi + \sqrt{\xi}\right),$$
$$\lambda_3 = \lambda_4 = \frac{27}{64\pi^2}\left(\psi - \sqrt{\xi}\right),$$

where

$$\psi = 2\pi S_{aa} + S_{bb},$$
$$\xi = 4\pi^2 S_{aa}^2 - 4\pi S_{aa} S_{bb} + S_{bb}^2 + 8\pi S_{ab}^2.$$

When $c = 1$, there are $m = 2$ nonzero eigenvalues given by

$$\lambda_1 = \lambda_2 = \frac{27}{32\pi^2}(2\pi + 1).$$

8.3. The permutation group

For p-dimensional data, let $\mathcal{G}$ be the group of order $g = p!$ consisting of the $p \times p$ permutation matrices. For this group, a fundamental region is

$$G = \{x \in \mathbb{R}^p : x_1 < x_2 < \cdots < x_p\}.$$

The $p!$ regions G_i then correspond to all the possible orderings of the data. Let $X_{(1)}, \ldots, X_{(p)}$ be the order statistics of a sample of size p from the $N(0, 1)$ distribution. In this situation, it is easily seen that

$$EX_{(i)} = p!\, v_i,$$

$$EX_{(i)}X_{(j)} = p!\, M_{ij}$$

Table 12
Nonzero eigenvalues of $Q^t Q$ for the permutation group

	Eigenvalue	Multiplicity
$p = 2$	$0.6366198\, S_{aa}$	1
$p = 3$	$0.7161972\, S_{aa}$	2
	$0.341959\, S_{bb}$	2
$p = 4$	$0.7652196\, S_{aa}$	3
	$0.4886807\, S_{bb}$	2
	$0.4052847\, S_{bb}$	3
$p = 5$	$0.7987651\, S_{aa}$	4
	$0.5536405\, S_{bb}$	5
	$0.4545718\, S_{bb}$	4

so that the entries of v and M can be obtained from tables of the means and covariances of normal order statistics such as those in Tietjen et al. [24]. Using these values, one can numerically compute the required eigenvalues of $Q^t Q$. We have done this for $p = 2, 3, 4, 5$, obtaining expressions valid for all c. The nonzero eigenvalues and their multiplicities are given in Table 12.

9. Products of groups

A convenient way to construct groups of orthogonal matrices on higher dimensional spaces is by taking products. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be groups of orthogonal matrices, with $\mathcal{G}_i$ consisting of g_i matrices $\{\Gamma^i_1, \Gamma^i_2, \ldots, \Gamma^i_{g_i}\}$ which are $p_i \times p_i$. Define $\mathcal{G}$ to be the group containing all the block diagonal matrices $\mathrm{diag}(\Gamma^1_j, \Gamma^2_k)$. This group has order $g = g_1 g_2$ and its matrices are $p \times p$ with $p = p_1 + p_2$. Let G^1 and G^2 be fundamental regions for the groups $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively. It is clear that the Cartesian product $G^1 \times G^2$ is a fundamental region for $\mathcal{G}$. The space $\mathbb{R}^p$ is divided into g regions of the form $G^1_j \times G^2_k$ where $G^1_j = \Gamma^1_j G^1$ and $G^2_k = \Gamma^2_k G^2$.

Our earlier theory now applies; for any integer $c > 0$, the statistic X^2 based on gc cells constructed using the group $\mathcal{G}$ will have the limiting distribution given in (2.5). We need only calculate the relevant eigenvalues λ_i. It turns out that if the eigenvalue calculations can be done for the groups $\mathcal{G}^1$ and $\mathcal{G}^2$, they can also be done for $\mathcal{G}$. We now show how to do this.

Suppose we have computed the matrices corresponding to $\tilde{A}, \tilde{B}, \tilde{C}$ in (7.4) for each of the groups $\mathcal{G}_1$ and $\mathcal{G}_2$. Call the resulting matrices $\tilde{A}^1$, $\tilde{B}^1$, $\tilde{C}^1$ and $\tilde{A}^2$, $\tilde{B}^2$, $\tilde{C}^2$, respectively. Now consider the integrals in (7.4) expressed in terms of the group $\mathcal{G}$. A little thought shows the following:

$$\int_{G^1_i \times G^2_j} x_k f_0(x)\, dx = \begin{cases} g_2^{-1} \tilde{A}^1_{ik} & \text{if } k \leqslant p_1, \\ g_1^{-1} \tilde{A}^2_{jk'} & \text{if } k > p_1, \end{cases}$$

$$\int_{G^1_i \times G^2_j} x_k^2 f_0(x)\, dx = \begin{cases} g_2^{-1} \tilde{B}^1_{ik} & \text{if } k \leqslant p_1, \\ g_1^{-1} \tilde{B}^2_{jk'} & \text{if } k > p_1, \end{cases}$$

$$\int_{G_i^1 \times G_j^2} x_k x_\ell f_0(x)\, dx = \begin{cases} g_2^{-1} \tilde{C}^1_{im} & \text{if } k, \ell \leqslant p_1, \\ g_1^{-1} \tilde{C}^2_{jm'} & \text{if } k, \ell > p_1, \\ \tilde{A}^1_{ik} \tilde{A}^2_{j\ell'} & \text{if } k \leqslant p_1, \ell > p_1, \end{cases}$$

where $k' = k - p_1$, $\ell' = \ell - p_1$, and m and m' are the positions of $x_k x_\ell$ and $x_{k'} x_{\ell'}$ in the p_1 and p_2-dimensional versions of $r(x)$, respectively. From these expressions it is clear that, with an appropriate ordering of the rows, the matrices $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ can be written in terms of Kronecker products. For $i = 1, 2$, let p^i denote the $g_i \times 1$ vector defined by $p^i = g_i^{-1} e_{g_i}$. Then

$$\begin{aligned} \tilde{A} &= (\tilde{A}^1 \otimes p^2, p^1 \otimes \tilde{A}^2), \\ \tilde{B} &= (\tilde{B}^1 \otimes p^2, p^1 \otimes \tilde{B}^2), \\ \tilde{C} &= (\tilde{C}^1 \otimes p^2, p^1 \otimes \tilde{C}^2, \tilde{A}^1 \otimes \tilde{A}^2), \end{aligned}$$

where in writing $\tilde{C}$ we have chosen a convenient ordering of the coordinates of $r(x)$ for $x \in \mathbb{R}^p$.

Substitute the above expressions into (7.6) to obtain Q. We note that the eigenvalues of $Q^t Q$ are not changed by permuting the columns of Q. By suitably rearranging the columns of Q (and doing a little algebraic simplification), we obtain the matrix

$$\bar{Q} = \left(\bar{Q}_1, \bar{Q}_2, \sqrt{gc}\, b \otimes \tilde{A}^1 \otimes \tilde{A}^2 \right),$$

where

$$\begin{aligned} \bar{Q}_1 &= \sqrt{gc}\, (a \otimes \tilde{A}^1 \otimes p^2, b \otimes \tilde{D}^1 \otimes p^2)\,, \\ \bar{Q}_2 &= \sqrt{gc}\, (a \otimes p^1 \otimes \tilde{A}^2, b \otimes p^1 \otimes \tilde{D}^2), \end{aligned}$$

and $\tilde{D}^i$ is defined as in (7.6), but with $\tilde{B}$, $\tilde{C}$, g replaced by $\tilde{B}^i$, $\tilde{C}^i$, g_i, respectively. The columns of $\bar{Q}_1$, $\bar{Q}_2$, and $b \otimes \tilde{A}^1 \otimes \tilde{A}^2$ are easily seen to be mutually orthogonal so that $\bar{Q}^t \bar{Q}$ has the block diagonal form

$$\bar{Q}^t \bar{Q} = \operatorname{diag}\left[\bar{Q}_1^t \bar{Q}_1\,,\ \bar{Q}_2^t \bar{Q}_2\,,\ S_{bb} \left(g_1 (\tilde{A}^1)^t \tilde{A}^1 \right) \otimes \left(g_2 (\tilde{A}^2)^t \tilde{A}^2 \right) \right]. \tag{9.1}$$

The matrices $\bar{Q}_i^t \bar{Q}_i$ turn out to have the same form as $Q^t Q$ in (7.7), but with g, $\tilde{A}$, $\tilde{D}$, replaced by g_i, $\tilde{A}^i$, $\tilde{D}^i$. (Note, however, that the quantities S_{aa}, S_{bb}, S_{ab} involved in $\bar{Q}_i^t \bar{Q}_i$ are those for dimension p, not p_i.)

We now know the eigenvalues of $Q^t Q$. Let $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}^1$, $\boldsymbol{\lambda}^2$, $\boldsymbol{\lambda}^1_A$, $\boldsymbol{\lambda}^2_A$ denote row vectors containing the nonzero eigenvalues of $Q^t Q$, $\bar{Q}_1^t \bar{Q}_1$, $\bar{Q}_2^t \bar{Q}_2$, $g_1 (\tilde{A}^1)^t \tilde{A}^1$, $g_2 (\tilde{A}^2)^t \tilde{A}^2$, respectively. From (9.1) we see that

$$\boldsymbol{\lambda} = (\boldsymbol{\lambda}^1, \boldsymbol{\lambda}^2, S_{bb}\, \boldsymbol{\lambda}^1_A \otimes \boldsymbol{\lambda}^2_A)\,.$$

If $(\tilde{A}^i)^t \tilde{D}^i = 0$ for $i = 1, 2$, then (as in (7.7)) the matrices $\bar{Q}_i^t \bar{Q}_i$ assume a block diagonal form so that (as in (7.8)) we have

$$\boldsymbol{\lambda}^i = (S_{aa} \boldsymbol{\lambda}^i_A, S_{bb} \boldsymbol{\lambda}^i_D),$$

where $\boldsymbol{\lambda}^i_D$ is a row vector of the nonzero eigenvalues of $g_i (\tilde{D}^i)^t \tilde{D}^i$. Thus, in this special case, the eigenvalues of $Q^t Q$ are

$$\boldsymbol{\lambda} = (S_{aa} \boldsymbol{\lambda}^1_A, S_{aa} \boldsymbol{\lambda}^2_A, S_{bb} \boldsymbol{\lambda}^1_D, S_{bb} \boldsymbol{\lambda}^2_D, S_{bb}\, \boldsymbol{\lambda}^1_A \otimes \boldsymbol{\lambda}^2_A)\,. \tag{9.2}$$

Table 13
Eigenvalues for a product of bivariate groups with $g_1, g_2 \geqslant 5$

Eigenvalue	Multiplicity
$\alpha_1 S_{aa}$	2
$\alpha_2 S_{aa}$	2
$\delta_1 S_{bb}$	2
$\delta_2 S_{bb}$	2
$\alpha_1 \alpha_2 S_{bb}$	4

It must be kept in mind that the quantities S_{aa}, S_{ab}, S_{bb} depend on both p and c. Thus, $\boldsymbol{\lambda}^1$ and $\boldsymbol{\lambda}^2$ above are not the same values obtained when solving for the eigenvalues in the p_1 and p_2-dimensional problems.

Example. To illustrate these results, we take the case where $\mathcal{G}_1$ and $\mathcal{G}_2$ are groups of 2×2 rotation matrices as discussed in Section 8.2. The product group $\mathcal{G}$ then consists of 4×4 matrices. So long as $g_1 \neq 3$ and $g_2 \neq 3$, we have $(\tilde{A}^i)^t \tilde{D}^i = 0$ for $i = 1, 2$ and (9.2) applies. In addition, assume for convenience that both g_1 and g_2 are at least 5 so that we may use the results in (8.2). The vectors $\boldsymbol{\lambda}_A^i$ and $\boldsymbol{\lambda}_D^i$ may be read from (8.2) (see (7.8)). If we let

$$\alpha_i = \frac{g_i^2}{8\pi}\left(1 - \cos\left(2\pi/g_i\right)\right) \quad \text{and} \quad \delta_i = \frac{g_i^2}{16\pi^2}\left(1 - \cos\left(4\pi/g_i\right)\right),$$

then we may write

$$\boldsymbol{\lambda}_A^i = (\alpha_i, \alpha_i) \quad \text{and} \quad \boldsymbol{\lambda}_D^i = (\delta_i, \delta_i).$$

Now (9.2) tells us that the eigenvalues of $Q^t Q$ are as given in Table 13. These eigenvalues determine the limiting distribution in (2.5) for the statistic X^2 based on $g = g_1 g_2$ sectors in $\mathbb{R}^4$ where the sectors are obtained as products of two-dimensional angular regions.

Acknowledgements

The authors thank the referees for comments and suggestions which have improved the paper.

References

[1] L. Baringhaus, N. Henze, A consistent test for multivariate normality based on the empirical characteristic function, Metrika 35 (1988) 339–348.
[2] L. Baringhaus, N. Henze, Limit distributions for Mardia's measure of multivariate skewness, Ann. Statist. 20 (1992) 1889–1902.
[3] R. Beran, Testing for ellipsoidal symmetry of a multivariate density, Ann. Statist. 7 (1979) 150–162.
[4] C.F. Dunkl, Y. Xu, Orthogonal Polynomials of Several Variables, Cambridge University Press, Cambridge, U.K., 2001.
[5] P.S. Dwyer, Some applications of matrix derivatives in multivariate analysis, J. Amer. Statist. Assoc. 62 (1967) 607–625.
[6] M.D. Ernst, A multivariate generalized Laplace distribution, Comput. Statist. 13 (1998) 227–232.
[7] K.T. Fang, T.W. Anderson, Statistical Inference in Elliptically Contoured and Related Distributions, Allerton Press, New York, 1990.
[8] K.T. Fang, S. Kotz, K.W. Ng, Symmetric multivariate and related distributions, Monographs on Statistics and Applied Probability, vol. 36, Chapman & Hall, London, 1990.

[9] L.C. Grove, C.T. Benson, Finite Reflection Groups, second ed., Springer, New York, 1985.
[10] N. Henze, T. Wagner, A new approach to the BHEP tests for multivariate normality, J. Multivariate Anal. 62 (1997) 1–23.
[11] D.J. Hodgson, K.P. Vorkink, Efficient estimation of conditional asset-pricing models, J. Business Econom. Statist. 21 (2003) 269–283.
[12] F.W. Huffer, C. Park, The limiting distribution of a test for multivariate structure, J. Statist. Plann. Inference 105 (2002) 417–431.
[13] J.P. Imhof, Computing the distribution of quadratic forms in normal variables, Biometrika 48 (1961) 419–426.
[15] V. Koltchinskii, L. Sakhanenko, Bootstrap testing for ellipsoidal symmetry of a multivariate distribution, in: E. Giné, D.M. Mason, J.A. Wellner (Eds.), High Dimensional Probability II, Progress in Probability, vol. 47, Birkhauser, Boston, 2000, pp. 451–493.
[16] K.C. Li, Sliced inverse regression for dimension reduction, J. Amer. Statist. Associat. 86 (1991) 316–327.
[17] A. Manzotti, F.J. Pérez, A.J. Quiroz, A statistic for testing the null hypothesis of elliptical symmetry, J. Multivariate Anal. 81 (2002) 274–285.
[18] A. Manzotti, A.J. Quiroz, Spherical harmonics in quadratic forms for testing multivariate normality, Test 10 (2001) 87–104.
[19] K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970) 519–530.
[20] D.S. Moore, J.B. Stubblebine, Chi-square tests for multivariate normality with application to common stock prices, Commun. Statist. Theor. Meth. 10 (1981) 713–738.
[21] D. Pollard, General chi-square goodness-of-fit tests with data-dependent cells, Z. Wahrsch. verw. Gebiete 50 (1979) 317–332.
[22] J.L. Romeu, A. Ozturk, A comparative study of goodness-of-fit tests for multivariate normality, J. Multivariate Anal. 46 (1993) 309–334.
[23] J.R. Schott, Testing for elliptical symmetry in covariance-matrix-based analyses, Statist. Probab. Lett. 60 (2002) 395–404.
[24] G.L. Tietjen, D.K. Kahaner, R.J. Beckman, Variances and covariances of the normal order statistics for sample sizes 2 to 50, Selected Tables in Mathematical Statistics, vol. 5, American Mathematical Society, Providence, RI, 1977.
[25] L.-X. Zhu, G. Neuhaus, Conditional tests for elliptical symmetry, J. Multivariate Anal. 84 (2003) 284–298.